

INFO-2030

Primljeno / Received: 2011-05-10

UDK: 004.032.26:519.237.8:657

Izvorni znanstveni rad / Original Scientific Paper

NEURONSKA MREŽNA KLASIFIKACIJA U MENADŽERSKOM RAČUNOVODSTVU

CLASSIFICATION WITH NEURAL NETWORKS IN MANAGERIAL ACCOUNTING

Brano Markić

Ekonomski fakultet, Sveučilište u Mostaru, Mostar, Bosna i Hercegovina
Faculty of Economics, University of Mostar, Mostar, Bosnia and Herzegovina

Sažetak

Menadžersko računovodstvo je orijentirano na generiranje informacija menadžerima koji unutar organizacije usmjeravaju i nadziru procese. Zato se mora pripremati veliki broj analiza i izvješća koji uspoređuju planirane i ostvarene rezultate ili izvješća koja sadrže indikatore o rezultatima poslovanje. Česti zadatak menadžerskom računovodstvu je klasifikacija. Problem klasifikacije se pojavljuje kada neki objekt treba pridružiti unaprijed definiranoj grupi ili klasi na temelju vrijednosti promatranih atributa. Tradicionalni pristup klasifikaciji je diskriminacijska analiza ili Bayesova teorija odlučivanja koja izračunava posteriori vjerojatnosti i na temelju njih donosi odluku o klasifikaciji. U radu se analizira i prikazuju rezultati klasifikacije na temelju neuronskih mreža i njihovog povezivanja, integracije s klaster analizom (k-means algoritam klasteriranja). Izgrađen je sustav CLUSTERNET koji iz skladišta podataka (datawarehouse) u skladu s zahtjevom menadžera generira željeni broj klastera a potom pomoću neuronske mreže izgrađenoj u jeziku R sve nove objekte klasificira u odgovarajuće klustere.

Abstract

Managerial accounting is oriented to producing information to managers – that is, people inside an organization who control and direct processes and operations. Therefore has to be prepared a large number of analysis and reports comparing actual results with plans or reports with many indicators about business results. Very often task in managerial accounting is classification. A classification problems occur when an object must to be assigned into predefined group or class based on the values of observed attributes. The traditional approach to classification is Bayesian theory of decision making. In the paper is analysed another approach based on neural networks and their linking and integration with cluster analyse (k-means algorithm). It was built CLUSTERNET system which from data warehouse according to manager's requirements produces a wanted number of clusters and using neural network, built in language R, classifies each one new object into corresponding cluster.

1. UVOD

Menadžersko računovodstvo je orijentirano analizi poslovanja i generiranju informacija potrebnih za donošenje odluka na različitim razinama organiziranosti unutar objektnog sustava. Ono je orijentirano odlučivanju čiji je ključni aspekt predvidjeti okolnosti povezane s pojedinačnom situacijom odlučivanja.

Različitost situacija odlučivanja jasno ilustrira da jedna metoda predviđanja, klasifikacije, planiranja ili neki njihov mali skup nisu dovoljni da zadovolje potrebe menadžmenta. Menadžersko računovodstvo nastoji osigurati potrebne

informacije za odlučivanje u različitim funkcijskim područjima poslovanja. Ono se služi podacima zabilježenim u bazama podataka i na temelju njih „otkriva“ skrivene veze i odnose koje postoje između pojedinih ekonomskih varijabli i veličina. Menadžersko računovodstvo nastoji predvidjeti budućnost i ne fokusira se samo na bilježenje poslovnih događaja iz prošlosti.

Složene zadatke menadžerskog računovodstva danas je moguće ispuniti jedino implementacijom odgovarajućih informatičkih koncepata i metoda a njihov razvoj menadžerskog računovodstva dodatno inspirira i postavlja im nove izazove. Baze podataka, skladišta podataka, upravljački informacijski

sustavi, sustavi za potporu odlučivanju, otkrivanje znanja u bazama podataka, data mining algoritmi, poslovna inteligencija, metode umjetne inteligencije, ekspertni sustavi, strojno učenje doprinose razvoju menadžerskog računovodstva i vice versa. Razvoj menadžerskog računovodstva ovisi o napretcima u pobrojanim informatičkim konceptima što transformira menadžersko računovodstvo u izrazito interdisciplinarno područje. Između tih koncepata i menadžerskog računovodstva snažna je integracija u našoj suvremenosti.

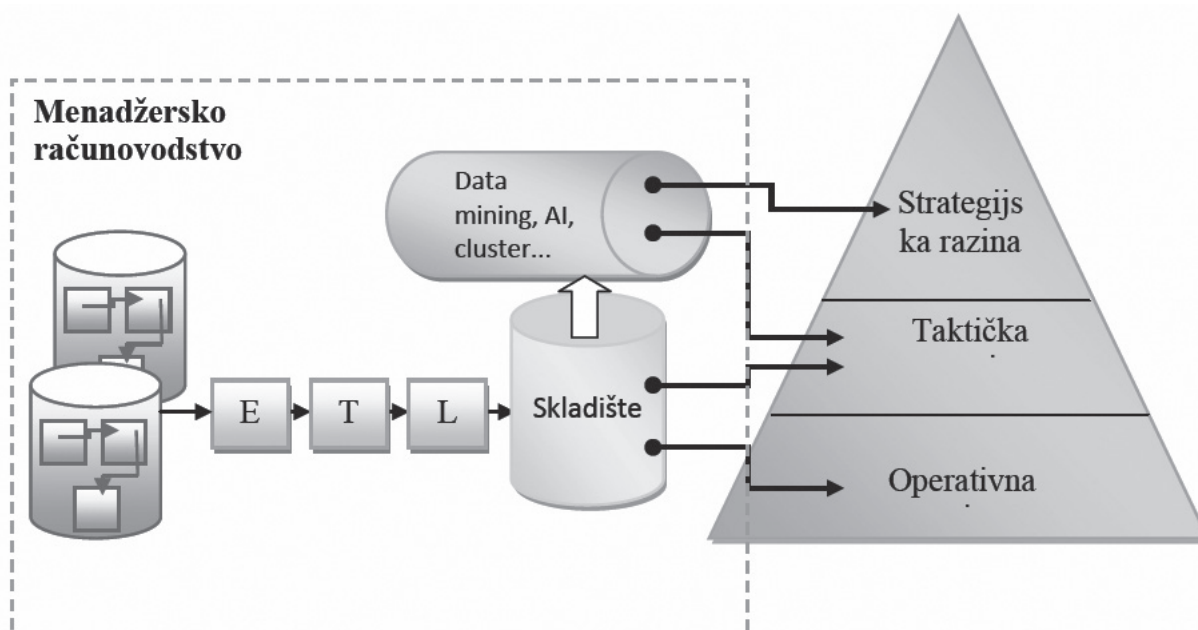
2. MENADŽERSKO RAČUNOVODSTVO I GENERIRANJE INFORMACIJA

Menadžersko računovodstvo proizvodi informacije za različite probleme odlučivanja i za različite razine upravljanja. Odluke na pojedinim razinama upravljanje donose se na temelju informacija a njih osigurava menadžersko računovodstvo. Veliki je broj mogućih problema odlučivanja i sve njih unaprijed nije moguće predvidjeti. Međutim, postoje zadaci i postupci koji su primjenjivi u različitim problemskim situacijama. Takvi su deskripcija, klasifikacija, klasteriranje, predviđanje. Oni su zato česti zadaci za menadžersko računovodstvo primjenjivi u gotovo svim poslovnim funkcijama. Primjera je mnogo. Neki od njih su klasifikacija troškova, potraživanja, obveza, imovine, dobavljača, kupaca. Predviđanje prihoda, troškova, obveza, potraživanja, razlike u cijeni, dobiti, cijena dionica. Tako je u marketingu posebno važno klasteriranje kupaca s obzirom na

njihova različita obilježja (demografska, socijalna, ekonomska). Određivanje klastera postupak je formiranja klasa na temelju podataka. Razlika između klasteriranja i klasifikacije je u tome što kod klasteriranja kriteriji klasifikacije nisu unaprijed poznati. Poznat je unaprijed samo broj klastera. Cilj je maksimizirati razlike između klastera i minimizirati udaljenost unutar klastera.

Menadžersko računovodstvo na temelju podataka iz financijskog računovodstva ekstrahira za svakog kupca obilježja na temelju kojih se oni svrstavaju u klaster. Između velikog broja atributa o kupcima A1, A2, ...Ak biraju se samo atributi relevantni za klasteriranje (to je podskup atributa A1, A2, ...Ak). Tako se osiguravaju financijske i ekonomske informacije za menadžere i druge korisnike unutar organizacije.

Podaci kojima se služi menadžersko računovodstvo nalaze se u bazama podataka (skupa međusobno povezanih tablica ili klasa). „Podaci u bazama podataka su dinamički. Mijenjaju se u vremenu za razliku od skladišta podataka koje predstavlja vremenski ovisne, ali nepromjenjive i predmetno orijentirane podatke.“ /1/ U dinamičkom smislu menadžersko računovodstvo sadrži niz međusobno povezanih i ovisnih aktivnosti. Prva je selekcija podataka, a zatim njihovo filtriranje i pročišćavanje (proces poznat kao ekstrakcija (E), transformacija (T) i punjenje (L). Rezultat selekcije su podaci koji se često moraju dodatno transformirati i obraditi jednostavnim agregiranjem. Takav sekvencijalan niz aktivnosti prikazan je kao E>T>L proces na sljedećoj slici.



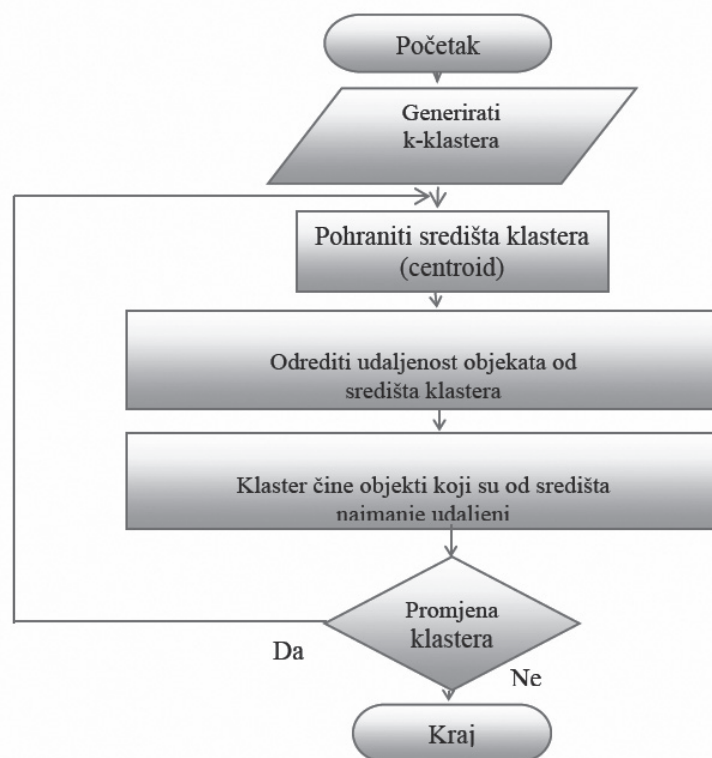
Slika 1.: Menadžersko računovodstvo i generiranje informacija za odlučivanje

„Prva aktivnost zove se *selekcija* (izbor) podataka. Rezultat selekcije ciljni su podaci koji se selektiraju iz skladišta podataka ili iz baza podataka. Ti su podaci najčešće u obliku tablica relacijskog modela. Selekcija se koncentrira na podskup podataka neke baze podataka, a taj podskup je bitan za proces generiranja informacija. Naravno, to nije jednostavna aktivnost“ /2/. Ona pretpostavlja jasno postavljene i definirane informacijske zahtjeve. Filtriranje, a nakon njega i pročišćavanje podataka, ima za cilj ukloniti potencijalnu inkonzistenciju (neažurnost) podataka, a na koje će se primijeniti algoritam za rudarenje podacima (data mining). Izbor algoritma ovisit će o informacijama potrebnim u poslovnom odlučivanju, o postavljenom zadatku. Postavljeni zadatak može biti klasifikacija, modeliranje ovisnosti, otkrivanje promjene ili odstupanja u odnosu na postavljeni cilj, određivanje klastera.[M]

2.1. Klasteriranje

Klasteriranje je postupak podjele nekog skupa podataka na unaprijed zadani broj klastera tako da je sličnost među elementima klastera najveća a razlika (udaljenost) među klasterima najveća. Postoje različiti tipovi klasteriranja. Tako npr. sakupljajuće (eng. *agglomerative*) hijerarhijsko klasteriranje u početku svakoj točki pridružuje jedan klaster. U svakom sljedećem koraku sakuplja najbliži par klastera (one koji su najmanje udaljeni) u novi klaster. Postupak sakupljanja se nastavlja sve dok ne ostane jedan (ili k) klastera.

Razdvajajuće (eng. *divisive*) počinje s klasterom koji sadrži sve točke. U sljedećim koracima klaster se dijeli sve dok svaki klaster postane samo jedna točka. U radu se primjenjuje k-means postupak klasteriranja /3/.



Slika 2.: Blok dijagram algoritma k-means

Algoritam k-means klasteriranja čini niz sljedećih koraka:

1. Slučajno odabrati k-klastera
2. Odrediti središta (centroide) za svaki klaster
3. Ponavljati:

Odrediti udaljenost objekata od središta klastera i objekt pridružiti najbližem klasteru.

Ponovno izračunati središta klastera.

Ponavljati sve dok objekti prelaze iz jednog u drugi klaster (uvjet završetka algoritma).

Podaci su smješteni u relacijskog tablici koja se ekstrahira iz transakcijske baze (financijskog računovodstva). Relacijska tablica je DataKlaster i čuva podatke o identifikacijskom broju (atribut Šifra), prometu s kupcem za godinu dana (atribut Promet), ostvarenoj razlici u cijeni (atribut RUC) i prosječnom broju dana plaćanja računa (atribut DOP) /4/.

Nakon primjene algoritma k-means objekti (kupci) se pridružuju jednom od četiri klastera a što ilustrira sljedeća tablica:

2.1.1 Podaci

Data				DataKlaster				
Šifra kupca	Promet	RUC	DOP	Šifra kupca	Promet	RUC	DOP	Klaster
1	234	24	30	1	234	24	30	1
2	23456	3451	45	2	23456	3451	45	4
3	1341	45	56	3	1341	45	56	1
4	1245	89	23	4	1245	89	23	1
5	123	23	45	5	123	23	45	1
6	5644	458	30	6	5644	458	30	1
7	14351	2340	55	7	14351	2340	55	4
8	755	70	65	8	755	70	65	1
...

Algoritam
k-means
(CLUSTERNE)

Tablica 1. Primjena

Služeći se programskim jezikom R /5/ moguće je vizualizirati rezultate primjene algoritma k-means u trodimenzijskom sustavu¹. Dovoljno je napisati sljedeći niz naredbi u jeziku R:

```
> library(RODBC)
> infer=odbcConnectAccess("C:\\Data.mdb") //
Učitava se baza podataka „Data.mdb“.
> inferP=sqlFetch(infer,"DataKlaster")
//Iz baze se bira tablica „DataKlaster“.
> Promet=inferP[,2] //Vektor2 Promet čuva podatke
```

o prometu kupca (drugi stupac tablice „DataKlaster“).
> RUC=inferP[,3] //Vektor RUC čuva podatke o razlici u cijeni.

> Klaster=inferP[,5] //Vektor Klaster čuva podatke o klasteru kupca (rezultat algoritma k-means).

Naredba:

```
> Prikaz3D=scatterplot3d(Promet, RUC, Klaster,
pch=16, highlight.3d=TRUE, type="h",
+ main="Prikaz klasteriranja u tri dimenzije")
vizualizira četiri klastera (slika 3.).
```

¹ Zanimarito ćemo dimenziju prosječan broj dana plaćanja kako bi mogli vizualizirati klaster (rezultate primjene k-means algoritma). Za prikaz u trodimenzijskom sustavu potrebno je instalirati Package *scatterplot3d* tako što na URL adresi <http://cran.r-project.org/web/packages/scatterplot3d> i datoteku *Windows binary:scatterplot3d 03-30.zip*. Potom pozvati u izborniku R-a opciju Packages i podizbornik **Install package(s) from local zip files...**

² Osnovni tip podataka u jeziku R je vektor. On je indeksirani skup vrijednosti a te vrijednosti su istoga tipa. Vektori mogu sadržavati numerički, znakovni i logički tip podataka. Nije moguće kombinirati vektore s različitim tipovima podataka. Naredbom *class(vektor)* moguće je provjeriti tip podataka vektora. U našem primjeru stupac *inferP[,3]* je vektor koji sadrži numerički tip podataka:
> *class(inferP[,3])*
[1] "numeric"

Vektor ima isti tip dok liste mogu sadržavati različite tipove podataka. Objekti liste se nazivaju komponente. Liste se najčešće kreiraju eksplicitno.

Primjer liste

```
L<-list(ime="Dijana", dob="27", broj.braće=3,
+ dob.braće=c(23,20,16), imena.braće=c("Marko", "Antonio", "Marin"))
```

Pristupiti članu (komponenti) liste se može pomoću indeksa. Tako je prvi član liste:

```
> L[[1]]
```

```
[1] "Dijana"
```

Peti član liste je vektor:

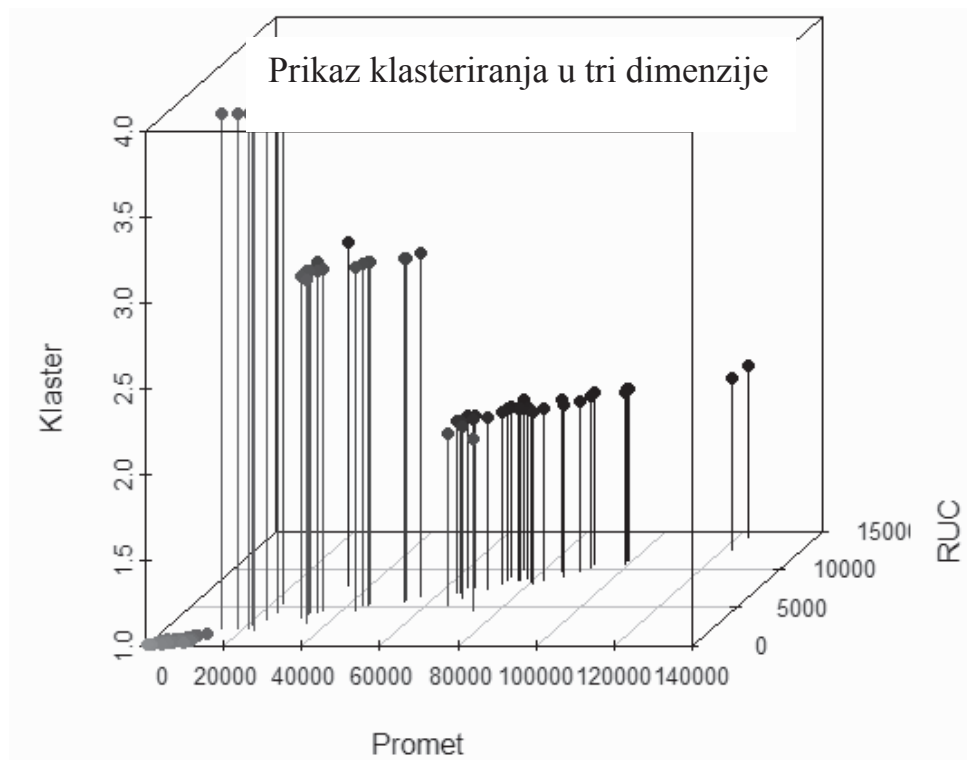
```
> L[[5]]
```

```
[1] "Marko" "Antonio" "Marin".
```

Prva komponenta vektora u listi L se prikazuje naredbom:

```
> L[[5]][1]
```

```
[1] "Marko".
```



Slika 3.: Vizualizacija rezultata klasteriranja objekata

Svaki objekt (kupac) je potpuno smješten u jedan i samo jedan klaster a što čini glavnu slabost k-means algoritma u odnosu na fuzzy klasteriranje. Naime, fuzzy klasteriranje se služi posebnom funkcijom pripadnosti objekta klasteru (broj između nula i jedan $[0,1]$). Vrijednost funkcije pripadnosti jedan označava pripadnost kupca (objekta) samo tom klasteru a 0 znači da kupac (objekt) ne pripada klasteru. Uporabom funkcije pripadnosti objekta klasteru dobiva se preciznija informacija o pripadnosti objekta klasterima.

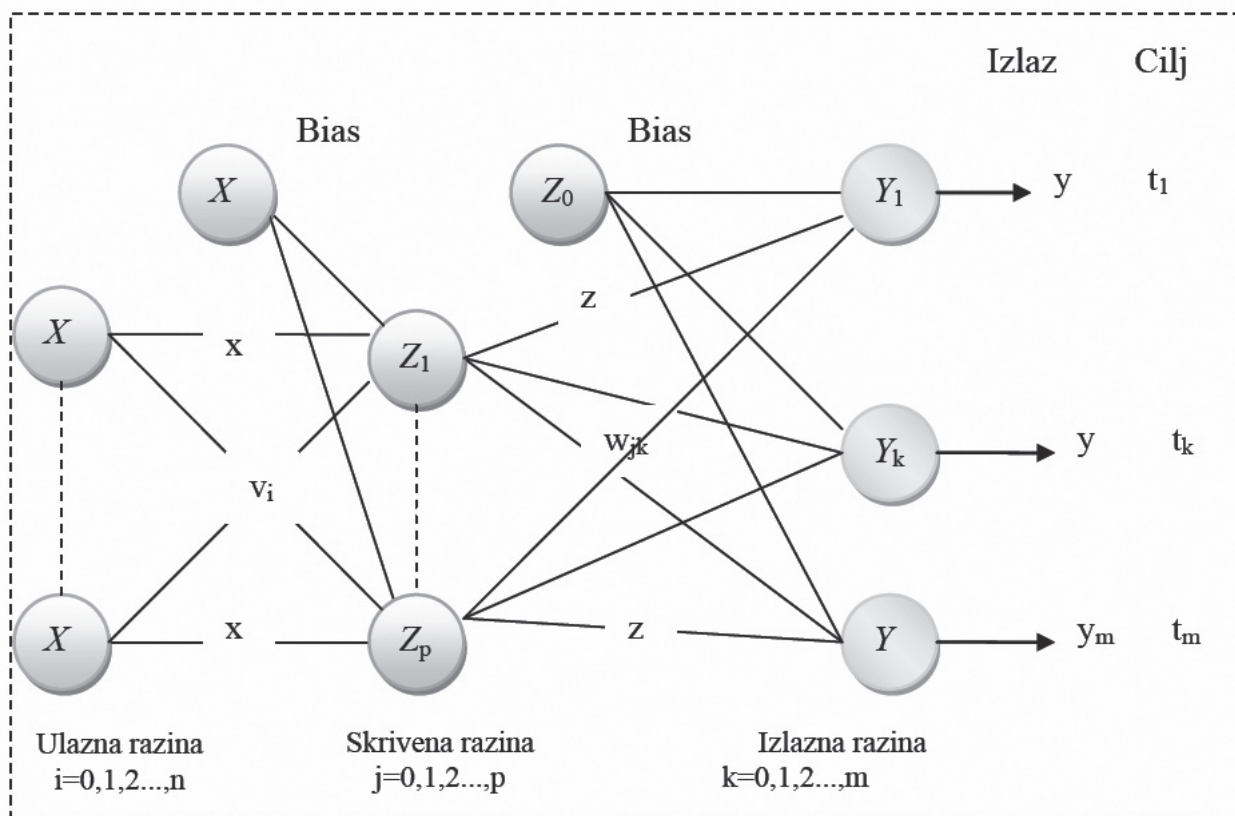
3. NEURONSKA MREŽA ZA PRIDRUŽIVANJE OBJEKATA KLASITERIMA

Neuronske mreže danas se koriste različitim oblastima. Neke od njih su ekonomija (npr. na finansijskim tržištima klasične metode su gotovo neprimjenjive jer se radi o nelinearnim funkcijama), ocjena kreditne sposobnosti komitenata banke kod odobravanja zajmova, vremenska prognoza, medicina (dijagnosticiranje bolesti, proučavanje rada mozga), prepoznavanje uzoraka.

Dijagram na sljedećoj slici prikazuje multi-layer perceptron neuronsku mrežu s jednom skrivenom razinom. Mreža ima tri sloja: ulazni, skriveni i

izlazni. Ulazna razina ima n neurona, p neurona je na skrivenoj razini, m neurona je na izlaznoj razini. Težinski faktori između ulazne i skrivene razine označeni su s v_{ij} , a težinski faktori između skrivene i izlazne razine su označeni s ω_{jk} , gdje je $0 \leq i \leq n$, $0 \leq j \leq p$, $0 \leq k \leq m$.³

³ Cybenko, Hornik (1989) su pokazali da neuronska mreža s jednom skrivenom razinom i sa sigmoidnom funkcijom aktivizacije može aproksimirati svaku funkciju sa željenom tačnošću (accuracy).



Slika 4.: Višeslojna neuronska mreža

Signal putuje od ulaznog k izlaznom sloju neuronske mreže. Cilj je za poznate vrijednosti ulaza minimizirati razliku između izlaznih i ciljnih vrijednosti. U našem primjeru cilj je za nove vrijednosti prometa, razlike u cijeni i prosječnog broja dana plaćanja računa nekog kupca (objekta) pridružiti tog kupca odgovarajućem objektu (jednom od četiri klastera).

3.1. Izgradnja neuronske mreže za klasifikaciju kupaca u jeziku R

R jezik je besplatan softver koji omogućuje analizu podataka ali istodobno i programski jezik u kojem se mogu kreirati funkcije a ima i sve logičke strukture programiranja, povezuje se s drugim programskim jezicima kao što su C++, Java, Perl itd. Za uporabu jezika R pretpostavka je da korisnik ima osnovna znanja iz statistike, ekonometrije, modeliranja i sl. Zanimljivost jezika R je to što korisnik može sam kreirati svoje pakete (package) i distribuirati ih drugim korisnicima.

U izgradnji neuronske mreže poslužiti ćemo se paketom *neuralnet*. Podaci su smješteni u bazi podataka „Data.mdb“ u korijenskoj mapi tvrdog diska. U bazi se nalazi datoteka „DataKlaster“ i

sljedećim nizom naredbi se u data frame *inferP* pohranjuju podaci iz datoteke „DataKlaster“:

	Šifra kupca	Promet	RUC	DOP	Klaster
1	1	234	24	30	1
2	2	23456	3451	45	4
3	3	1341	45	56	1
4	4	1245	89	23	1
5	5	123	23	45	1
6	6	5644	458	30	1

Tablica 2.: Podaci iz data frame *inferP*

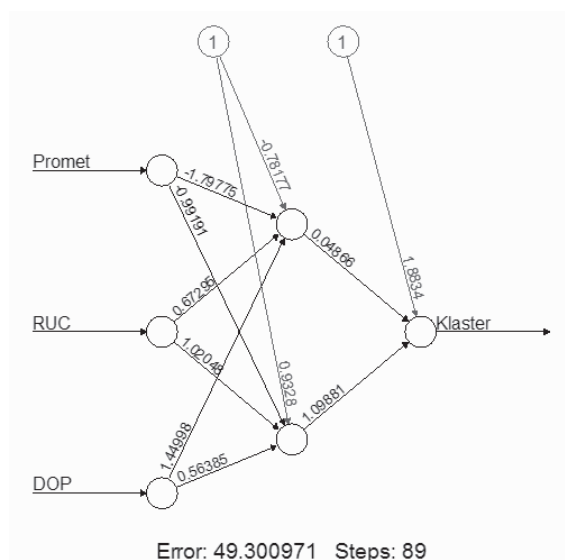
```
> library(RODBC)
> infer=odbcConnectAccess("C:\\Data.mdb")
> inferP=sqlFetch(infer,"DataKlaster")
> inferP
```

Potom se učitava paket *neuralnet*:

```
> library(neuralnet)
Loading required package: grid
Loading required package: MASS
```

Sljedeća naredba izgrađuje neuronsku mrežu:
nn.backpropagation<-neuralnet(Klaster~Promet
+RUC+DOP,inferP,hidden=2, learningrate=0.001,
algorithm="backprop", err.fct = "sse",linear.

output=TRUE)
Naredbom > plot(nn.backpropagation) crta
se odgovarajuća neuronska mreža s težinskim
faktorima:



Slika 5.: Neuronska mreža za pridruživanje kupca klasterima

Naredbom: n.izlaz=compute(nn.backpropagation
,covariate=matrix(c(32149,5440,46,2317,549,46),by
row=TRUE,ncol=3)) vrijednosti prometa razlike u
cijeni i prosječan broj dana plaćanja za nekog kupca
se pridružuju odgovarajućem klasteru.
> round(n.izlaz\$net.result,0)

```
[,1]
[1,] 2
[2,] 2
Dva objekta (prvi i drugi redak)
[,1] [,2] [,3] [,4]
[1,] 1 32149 5440 46
[2,] 1 2317 549 46
```

se pridružuju drugom klasteru. Tako je moguće svaki
novi objekt (kupac) s odgovarajućim vrijednostima
prometa, razlike u cijeni i prosječnog broja dana
plaćanja računa pridružiti jednom klasteru.

ZAKLJUČAK

Menadžersko računovodstvo generira informacije
za potrebe menadžera unutar organizacijskog
sustava. Njegov su fokus različite razine
menadžmenta unutar organizacije. Generiranje
potrebnih informacija za odlučivanje složen je
zadatak i menadžersko računovodstvo ga rješava
povezivanjem svojih znanja s znanjima drugih
disciplina. U radu je prikazan sustav CLUSTERNET
koji iz baza podataka puni skladište podataka a
potom podatke iz skladišta podataka klasterira
u četiri klastera uporabom algoritma k-means.
Na temelju rezultata klaster analize izgrađuje se i
trenira neuronska mreže u jeziku R koji pokazao
zadovoljavajuću aplikacijsku i razvojnu moć
uporabom paketa neuralnet.

Bilješke

- /1/ John Maindonald and John Braun, (2003) Data Analysis and Graphics Using R - An Example-Based Approach, Cambridge University Press.
- /2/ Ibidem
- /3/ Kukić, S., Markić, B., Metodologija društvenih znanosti – metode, tehnike, postupci i instrumenti znanstvenoistraživačkog rada, Ekonomski Fakultet Sveučilište u Mostaru, Mostar 2006.
- /4/ Markić, B., (2010) Marketing Intelligent Systems Using Soft Computing Studies in Fuzziness and Soft Computing, Volume 258/2010, Springer-Verlag Berlin Heidelberg 2010, pp 79-111.
- /5/ John Maindonald and John Braun, (2003) Data Analysis

and Graphics Using R - An Example-Based Approach, Cambridge University Press.

Literatura

1. De Vidal, Douglas and Ronald W. Hilton. (2008) Study guide for use with managerial accounting. New York: McGraw-Hill/Irwin, 2008. Type: Study Guide.
2. Hilton, Ronald W.. (2008) Managerial Accounting. 7th edition. New York: McGraw-Hill/Irwin, Type: Textbook. ISBN: 978-0-07-302285-7.
3. Markić, B., Tomić, D., (2006) Softverska rješenja u marketing istraživanju za otkrivanje znanja u bazama podataka pomoću fuzzy klasteriranja, INFORMATOLOGIJA, Vol.39 No.4 Prosinac 2006.